DOCUMENT RESUME

ED 268 175 TM 860 238

AUTHOR Albanese, Mark A.; And Others
TITLE An Observational Study of the Lecture Delivery Style Characteristics of High and Low Rated Lectures.
PUB DATE Apr 86
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
AVAILABLE FROM Office of Consultation and Research in Medical Education, College of Medicine, University of Iowa, Iowa City, IA 52242.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Analysis of Variance; *Behavior Rating Scales; Higher Education; Interrater Reliability; *Lecture Method; *Lesson Observation Criteria; *Medical School Faculty; *Student Evaluation of Teacher Performance; *Summative Evaluation; Teacher Effectiveness; Test Reliability
IDENTIFIERS *Lecture Characteristics Scale; Lecturers

ABSTRACT

This study identifies distinguishing differences in lecture delivery styles of lecturers rated by students in a large multi-instructor course: the Introduction to Clinical Medicine Course (ICM). The 20 lowest- and highest-rated lecturers of the 1982 and 1983 ICM courses served as the target group. Non-student raters observing the 1984 lectures completed the 49-item 6-option Lecture Characteristics Scale (LCS). The reliability of consensus ratings was greater than .70, indicating sufficient reliability to be used in subsequent analyses. To examine the relationship of student ratings of lecturers to observer LCS scores, t tests were completed for LCS subscores between the high and low rated lectures. For all six LCS subscores, statistically significant differences between the high and low lecturers were obtained. Out of 49 LCS items, 31 showed differences in means between the low and high groups that were statistically significant. Conclusions were: (1) student ratings of the most and least positively rated lectures showed a fair amount of stability across the three years studied; (2) independent observers were able to discriminate between lectures which students rated the highest and lowest; and (3) characteristics that discriminated the most effectively between the high and low rated lectures were the qualities associated with introduction/organization and student involvement. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED268175

AN OBSERVATIONAL STUDY OF THE LECTURE DELIVERY STYLE CHARACTERISTICS OF HIGH AND LOW RATED LECTURES

Mark A. Albanese, Ph.D.*
University of Iowa
College of Medicine
(319) 353-6781

Abstract

The primary objective of this study was to identify the distinguishing differences in lecture delivery styles of lecturers rated the highest and lowest by students in a large multi-instructor course (the Introduction to Clinical Medicine Course - ICM). Lecturers who gave lectures in ICM in 1982 and 1983 and who were also to give a lecture in the 1984 ICM Course were ranked from lowest to highest according to the mean of the overall rating by students in 1982 and 1983. The 20 lowest rated lecturers and the 20 highest rated lecturers, served as the target group. Non-student raters observed the lectures in 1984 (2 per lecture). The observers completed the 49 item 6-option Lecture Characteristics Scale (LCS). The reliability of concensus ratings were >.70 indicating sufficient reliability to satisfactorily be used in subsequent analyses. To examine the relationship of student ratings of lecturers to observer LCS scores, t tests were computed for LCS subscores between the high and low rated lectures. For all six LCS subscores, statistically significant differences between the high and low lectures were obtained. Out of 49 LCS items, 31 showed differences in means between the low and high groups that were statistically significant at the .05 level. It was concluded that: student ratings of the most positively and least positively rated lectures showed a fair amount of stability across the three years studied, independent observers were able to discriminate between lectures which students rated the highest and lowest, and characteristics that discriminated the most effectively between the high and low rated lectures were the qualities associated with the introduction/organization and student involvement.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

M. A. Albanese

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 860 238

*Sandra S. Matthes, M.A.T., University of Iowa College of Medicine, Iowa City, IA 52242; Diane E. Case, M.A., University of Iowa College of Medicine, Iowa City, IA 52242; Donald D. Brown, M.D., University of Iowa Hospitals and Clinics, Iowa City, IA 52242.

Key Words: Lecture, ratings, evaluation, reliability.

AN OBSERVATIONAL STUDY OF THE LECTURE DELIVERY STYLE CHARACTERISTICS OF HIGH AND LOW RATED LECTURES

Mark A. Albanese, Ph.D., University of Iowa College of Medicine; Sandra S. Matthes, M.A.T., University of Iowa College of Medicine, Iowa City, IA 52242; Diane E. Case, M.A., University of Iowa College of Medicine, Iowa City, IA 52242; Donald D. Brown, M.D., University of Iowa Hospitals and Clinics, Iowa City, IA 52242.

Lecture is one of the most common teaching methods used in Colleges at both the undergraduate and graduate levels. Student ratings of such lectures are often used by promotions committees in making decisions regarding faculty promotion and tenure. Although, student ratings are heavily used, there has been skepticism expressed as to whether they reflect the actual instruction received or other characteristics such as instructors' sex and rank, and student expected grades (Branderberg, D.C.; Slinde, J.A.; and Batista, E.E.; 1977). Therefore, it would be useful to determine to what degree student ratings actually relate to aspects of the lecture delivery style.

While extraneous factors like those mentioned might have significant influences on student ratings, they would be less likely to affect the ratings of non-student observers who are trained to focus on the delivery style characteristics of the lecturer. It follows that a study comparing the ratings of such individuals with the ratings provided by students would yield important information. However for such a study to be of most use, a variety of lecture delivery styles would need to be observed. The typical single lecturer course would not lend itself well to such a study as it would be difficult for a single individual to adopt a variety of delivery styles.

Request for reprints: Office of Consultation and Research in Medical Education, College of Medicine, University of Iowa, Iowa City, IA 52242, (319) 353-6781.

A multidisciplinary course involving large numbers of different lecturers would seem to be a more reasonable vehicle for the study.

Delivery style represents just one of perhaps many different instructional variables that might be studied. For instance, information density (Russell and Hendrickson, 1984) or structuring of information are additional variables that could have a major bearing on student ratings. Thus, a study of lecture delivery style is only examining part of the educational equation and might not be expected to show a very large effect size.

A procedure that would be likely to be sensitive to a fairly small relationship would be to contrast the extreme groups, i.e. compare the most favorably rated lecturers with the least favorably rated lecturers. Such a procedure would have some elements in common with Flanagan's critical incidents procedure (Flanagan, 1949). In Flanagan's procedure, successful and unsuccessful completions of a task are compared in great depth to determine the features which led to success or failure. Such a procedure has the potential to lead to specific recommendations for improved outcomes. While the specific criteria that would designate a lecture as a failure or success are open for debate, a contrasting of the most favorably rated lectures with the least favorably rated lectures in a given setting would seem likely to provide many of the same benefits.

This study was undertaken to explore to what degree lecture delivery style ratings provided by non-student observers in a large multi-instructor course would agree with ratings provided by students and to determine what lecture delivery style characteristics effectively discriminate between lectures students rate most and least favorably.

Methods

The Introduction to Clinical Medicine Course (ICM) at the University of Iowa College of Medicine has over 300 lectures given by approximately 200 lecturers. Each of the lectures in ICM is evaluated by a subgroup of 20-30 students. The rating form used by students to evaluate lectures consists of nine likert type questions plus an overall quality rating. The evaluation system has been in place for over six years and student response rates have ranged between 86-93% (Albanese, Schroeder and Barnes, 1979; Albanese and Matthes, 1983). The study plan called for carefully trained non-students to evaluate lecturer delivery styles using both quantitative and qualitative measurement instruments. For reasons stated earlier, the 20 most positively rated lecturers and 20 of the least positively rated lecturers were targeted. Since these lecturers had to be identified in advance, lecturers who made presentations in ICM in 1982 and 1983 and were also to give a lecture in the 1984 ICM Course were identified. These lecturers were ranked from lowest to highest according to the mean of the overall rating by students in the 1982 and 1983 ICM assessments (two years results were averaged to minimize regression effects). The 20 lecturers with the lowest mean (<3.40) and the 20 lecturers with the highest mean (>3.75) served as the target group.

The training involved about 10 hours of instruction in the recording of events occurring in a lecture situation using the form developed for the project. Also, the rating system was pilot tested in actual lectures in which the principal investigator compared notes and ratings with the observers. Once adequate agreement was achieved, the study observation began. Each lecture was evaluated by two observers to assess interrater reliability. Two separate pairs of raters were used in the study. After the lecture was over, the raters then met to discuss their individual ratings.

Each question for which the raters were more than one scale point apart was discussed until consensus was reached and noted in red pen on the evaluation form. If no consensus was able to be reached, during the analysis, the mean of the two raters was used as the question value. The observers were not told how lectures were selected in order to avoid biasing their findings.

The observers used two forms: the Lecture Characteristics Scale (LCS) and an Observation Form. The LCS consists of 49 6-option likert-type items from which six subscores are derived: Introduction/Organization (13 items), Voice Presentation Style (7 items), Non-verbal Presentation Style (6 items), Clarity of Presentation (5 items), Use of Audiovisual Aids (7 items) and Student Involvement (11 items). Subscores are means for the items in the section and therefore range from 1 to 6 (high scores were more positive). The observers completed this form independent of one another at the conclusion of the lecture.

The observation form was used by the observers to record notes as the lecture occurred. It has two sections labeled narrative and cues. The narrative comments section allowed the observer discretion to comment on all events occurring in the lecture. The cues section required comments on specific aspects of the instructional experience (e.g., eye contact, verbal characteristics, etc.). Observers were requested to comment on each of these specific aspects for each lecture to ensure observers were taking these characteristics into consideration as they assessed the lecture.

Results

Although the original plan called for 20 lectures in both the upper and lower groups, ad hoc changes in lecture scheduling resulted in one of the lower group lectures being missed. Thus, only 19 lectures were observed for the low group.

To what degree did the procedures used to select lectures succeed in identifying two extreme groups of lectures? The correlation of 1984 results with the 1982-1983 mean score (i.e., those scores upon which faculty were chosen to participate in the study) was .59. Overall, 65.8% (25 of 38) of the lecturers maintained their original classification when the original group selection criteria (low group <3.40, high group >3.75) were applied to the 1984 results. This compares with a 33% expectation by chance alone. There were two lecturers originally chosen for the upper group that did not meet the upper group criterion based on the 1984 results; one of which met the criterion for the lower group. There were 11 lecturers that changed from being in the lower group based on the 1982-83 composite to failing to meet this criterion in the 1984 results. Of the eleven, eight met the criterion for inclusion in the high group based on 1984 results. Thus, eight of 19 lecturers that formed the lower group based on the 1982-83 composite still met this criterion in the 1984 results while 17 of 19 lecturers that formed the upper group based on the 1982-83 composite continued to meet the selection criterion in the 1984 results. It is interesting to note that very few highly rated lecturers regressed and over half the negatively rated lecturers made substantial progress in the 1984 ratings.

The interrater reliabilities (Pearson product moment correlations) and internal consistency reliabilities (Alpha) of concensus scores are shown in Table 1.

(Insert Table 1 About Here)

Interrater reliabilities for five of the six LCS subscores exceeded .70 indicating satisfactory agreement for a study of this nature. However, the interrater reliability of the Clarity of Presentation subscore was .58 indicating unsatisfactory agreement among the raters. In spite of this

disagreement, the consensus ratings showed a comparatively high degree of internal consistency reliability (.72). This suggests that the disagreements were generally resolved in the discussions that followed the completion of the form. The observers themselves indicated that this was the case as well. Thus, while the reliability of the initial ratings by the observers for the subscore was inadequate, the consensus ratings showed sufficient reliability to satisfactorily be used in subsequent analyses.

To examine the relationship of student ratings of lecturers to observer LCS scores, t tests were computed for LCS subscores between the high and low rated lectures for the three different approaches to splitting the lectures.[1] First, the original criterion applied to the 1984 results. Next, the highest and lowest 15 lectures based on 1984 results and finally only those lecturers that met the criterion for each group all three years. The results of this analysis are shown in Table 2.

(Insert Table 2 About Here)

For all six LCS subscores, statistically significant differences between the high and low lectures were obtained for all three splits. To obtain a more meaningful understanding of what these subscore differences meant, t tests were computed for each item in the LCS for the split based on the highest and lowest 15. Table 3 contains the results of this analysis.

(Insert Table 3 About Here)

Out of 49 items, 31 showed differences in means between the high and low groups that were statistically significant beyond the .05 level. The upper

---

[1]The data were also analyzed via two factor analysis of variance with rater pair entered as the second factor. In no case was the rater pair or rater pair by group interaction statistically significant at the .05 level. Thus, the t tests were reported for simplicity of presentation.

bound on a 95% confidence interval for the number of significant differences one would expect by chance was 6. One might be inclined to attribute the large number of significant differences to "halo effects" (the tendency of raters to give similar ratings to distinctly different aspects of a lecture based on an overall perception of good or bad). However, there are a number of characteristics of the results that would argue against a halo effect. First, the means and standard deviations were comparatively variable across items comprising the subscore for the same group. The range of the means was at least .44 of a scale point (Voice presentation style—high group) and at most 2.83 (Use of audiovisuals—low group). The latter value covered over 50% of the score range suggesting a great deal of variability in responses from item to item. This is further reinforced by the comparatively large range in the item standard deviations on the various subtests. The minimum range was 0.41 scale points (voice presentation style—high group) while the largest range was 2.14 (student involvement—high group). Such variability in ratings would not be expected if the halo effect were the primary factor accounting for the results.

A comparison of the two groups based on comments made by the observers suggests that the lectures in the highly rated groups were more likely to: 1) maintain good eye contact by looking at the entire class (60% vs. 32%), 2) avoid a monotone presentation by using voice fluctuations (94% vs. 64%), 3) make good use of gestures and physical movement (49% vs. 23%). Low rated lecturers on the other hand were more likely to be difficult to hear because of speaking too softly, mumbling or a speech impairment. (25% vs. 1% high). A final lecture behavior worth mentioning is the repeating of student questions. Altough it was only done by two lecturers, both were in the high group. This behavior may be especially important to use in large lecture halls where it may be difficult for all students to hear a students questions.

Discussion

These results are encouraging in their implications for the use of student ratings in the evaluation of lectures. Lecturers were assigned to high and low groups on the basis of consistent ratings for a two-year period in order to minimize the impact of regression effects on the results. When stability of classification was examined, 66% of the lecturers maintained their original classification in 1984. This compares with 33% expectation through random assignment (since there is a possibility of three classifications--same group, middle, other extreme group). The fact that the great majority of changes constituted improvement from the lowest group, suggests that lecturers respond to student feedback by improving their presentation. This finding is consistent with those of Stillman, et al. (1983). In that study, feedback to lecturers was usually provided within 24 hours after the lecture; however, in the present study, feedback was delayed by up to 8 weeks to accommodate the processing of the data. Thus, even delayed feedback appears to assist lecturers in improving their presentations.

This study also showed that non-medical non-student observers were able to discriminate between those lectures students rate high and low. Those differences spanned the range of various lecture characteristics with 31 of 49 specific characteristics that were rated obtaining statistical significance. This suggests that such individuals might be useful in providing feedback in pre-presentation practice lectures. Regarding specific lecture delivery style characteristics that distinguished the high and low groups of lecturers, the observers noted introduction/organization qualities and student involvement efforts as being the most discriminating.

Conclusions

1. Student ratings of the most positively and least positively rated lectures showed a fair amount of stability across the three years studied.
2. Independent observers were able to discriminate between lectures which students rated the highest and lowest.
3. The characteristics that discriminated the most effectively between the high and low rated lectures were the qualities associated with the introduction/organization and student involvement.

References

Albanese, M.A.; Schoeder, J. and Barnes, H.V., "A Method of Assessing Student Perceptions of Instruction in a Multiple Instructor Course Which Maximizes Student Response." Evaluation and the Health Professions, 2(2):231-239, 1979.

Albanese, M.A.; Matthes, S. "Multiple Matrix Sampling As a Method to Improve Questionnaire Response Rates." Paper presented at the Joint Conference of the Evaluation Network and the Evaluation Research Society, Chicago, Illinois, Oct. 20-22, 1983.

Flanagan, J.C., "Critical Requirements: A New Approach to Employee Evaluation. Personnel Psychology, 2:419-425, 1949.

Irby, D.M., "Clinical Teacher Effectiveness in Medicine." Journal of Medical Education, 53:808-815, 1978.

Russell, I.J. and Hendricson, W.D., "Effects of Lecture Information Density a Medical Student Achievement," Journal of Medical Education, 59:881-889, 1984.

Stillman, P.; Gillers, M.; Heins, M.; Nicholson, G.; and Sabers, D.; "Effect of Immediate Student Evaluations On a Multi-Instructor Course," Journal of Medical Education, 58:172-178, 1983.

Table 1

Interrater Reliability and Internal Consistency Reliability (Alpha) for LCS Subscores

LCS Subscore	# of Items	Interrater Reliability[1]	Internal Consistency Reliability[2]
Introduction/Organization	13	.83	.83
Voice Presentation	7	.80	.75
Non-verbal Presentation	6	.89	.82
Clarity of Presentation	5	.58	.72
Use of Audiovisual Aids	7	.73	.97
Student Involvement	11	.93	.96
Total	49	.89	.93

[1] These estimates were obtained from two pairs of raters. To compute these values, the correlations for each rater pair were transformed by Fisher's Z, a weighted average computed and then reconverted to correlations. These values were then entered into the Spearman-Brown formula to arrive at an estimate for the mean of two raters.

[2] Estimates are based on the consensus data.

Table 2
Means and Standard Deviations of LCS Subscores
For Lecturers Receiving High and Low Student
Ratings Based on Three Different Splits

	N	Subscore: Introduction/ Organization (13 items)	Voice Presentation (7 items)	Non-Verbal Presentation (6 items)	Clarity of Presentation (5 items)	Use of AV Aids (7 items)	Student Involvement (11 items)	Total (47 items)
I. Split based on 1982-83 criterion applied to 1984 results (High < 2.25, Low ≥ 2.60)								
High	29	4.61 (0.57)[1]	5.03 (0.42)	5.01 (0.59)	4.70 (0.48)	4.72 (0.58)	4.04 (0.87)	4.69 (0.39)
Low	9	3.90 (0.46)	4.38 (0.54)	4.04 (0.76)	4.21 (0.52)	4.24 (0.28)	3.07 (1.11)	3.96 (0.40)
t		3.41**	3.78***	4.07***	2.59*	3.35**	2.72**	4.87****
II. Split based on 15 highest and 15 lowest 1984 scores (High < 1.85, Low ≥ 2.25)								
High	15	4.78 (0.62)	5.29 (0.35)	5.21 (0.55)	4.96 (0.26)	4.99 (0.40)	4.42 (0.75)	4.93 (0.28)
Low	15	4.00 (0.51)	4.44 (0.47)	4.24 (0.68)	4.21 (0.48)	4.26 (0.44)	3.34 (0.95)	4.07 (0.35)
t		3.76***	5.61****	4.26***	5.34****	4.78****	3.44**	7.38****
III. Split based on lecturers meeting original criterion and maintaining it in 1984 (High < 2.25, Low ≥ 2.6)								
High	15	4.78 (0.52)	5.10 (0.49)	4.91 (0.72)	4.77 (0.49)	4.87 (0.64)	4.10 (0.96)	4.79 (0.42)
Low	8	3.77 (0.28)	4.34 (0.56)	3.96 (0.77)	4.16 (0.53)	4.19 (0.25)	2.84 (0.92)	3.86 (0.27)
t		5.09****	3.33**	2.97**	2.77*	3.61**	3.04**	5.64****

[1]Values in parenthesis are standard deviations.

† $p < .10$
* $p < .05$
** $p < .01$
*** $p < .001$
**** $p < .0001$

Table 3
Item Means for the 15 Highest and 15 Lowest Rated

	High Mean	(SD)	Low Mean	(SD)	t
INTRODUCTION/ORGANIZATION					
1. Stated purpose of lecture	4.63	(1.14)	4.17	(1.38)	1.00
2. Indicated relevance of content	4.13	(1.51)	3.73	(1.61)	0.70
3. Outlined content	4.67	(1.26)	3.80	(1.50)	1.71†
4. Specified/defined instructional objectives	3.17	(1.76)	2.30	(1.44)	1.48
5. Arranged and discussed the lecture content in a systematic and organized fashion	5.50	(0.38)	4.67	(0.62)	4.46****
6. Related content to introductory purpose	5.21	(0.51)	4.77	(0.56)	2.16*
7. Made clear transitions	4.83	(0.59)	4.17	(0.67)	2.89**
8. Presented information at an appropriate level of "abstractness"	5.37	(0.35)	4.73	(0.50)	4.04***
9. Presented examples to clarify very abstract and difficult ideas.	5.32	(0.46)	4.57	(0.96)	2.72*
10. Defined terminology	5.03	(0.90)	3.82	(1.35)	2.86**
11. Emphasized important points/main ideas.	5.30	(0.59)	4.47	(0.72)	3.47**
12. Periodically summarized the most important ideas in the lecture	3.77	(1.39)	2.73	(1.46)	1.99†
13. Suggested ways to apply content	5.37	(0.44)	4.23	(1.55)	2.73*
PRESENTATIONS STYLE					
Voice					
1. Spoke at appropriate volume	5.13	(0.64)	4.13	(0.93)	3.42**
2. Raised or lowered voice for variety	5.17	(0.49)	4.17	(0.72)	4.44****

	High Mean	(SD)	Low Mean	(SD)	t
PRESENTATIONS STYLE(Voice cont.)					
3. Used voice for emphasis	5.27	(0.59)	4.23	(0.86)	3.82***
4. Avoided the use of speech fillers. ("okay", ahmm, etc.)	5.23	(0.73)	4.70	(0.90)	1.78†
5. Spoke neither too fast nor too slow	5.20	(0.62)	4.70	(0.82)	1.88†
6. Explained clearly and to the point	5.43	(0.32)	4.87	(0.52)	3.61**
7. Talked to class not to board or windows, etc.	5.57	(0.62)	4.27	(1.07)	4.08***
Non-verbal					
1. Used eye contact with the class	5.53	(0.64)	4.27	(1.00)	4.14***
2. Appeared natural--neither too stiff or too casual	5.43	(0.42)	4.73	(0.70)	3.31**
3. Seemed enthusiastic and interested in topic	5.43	(0.59)	4.47	(1.16)	2.88**
4. Moved purposefully--not pacing nor stuck to podium	4.53	(0.90)	3.47	(1.26)	2.67*
5. Used hands and arms for emphasis	4.83	(1.13)	3.70	(1.51)	2.33*
6. Facial and body movements did not contradict speech or expressed intentions. (Waited for response after asking a question.)	5.47	(0.48)	4.83	(0.31)	4.29***
CLARITY OF PRESENTATION					
1. Related new ideas to already familiar ones	4.93	(0.27)	4.47	(1.03)	1.68
2. Provided occasional summaries and restatements of important ideas	4.63	(0.85)	3.17	(1.21)	3.84***
3. Used alternate explanations when necessary	5.25	(0.42)	4.00	(0.89)	3.18**
4. Slowed word flow when ideas were complex and difficult	5.00	(0.53)	4.20	(0.88)	3.00**

	High Mean	(SD)	Low Mean	(SD)	t
CLARITY OF PRESENTATION(CONT.)					
5. Did not digress from main topic	5.13	(0.64)	5.07	(0.37)	0.35
AUDIOVISUAL AIDS					
1. Used microphone effectively	5.00	(0.71	4.57	(0.59)	1.79†
2. Used audiovisuals to enhance the verbal presentation.	5.40	(0.43)	4.77	(0.62)	3.24**
3. Coordinated audiovisual with verbal presentation	5.43	(0.37)	4.77	(0.53)	3.99***
4. Used well designed audiovisuals	5.27	(0.56)	4.20	(0.92)	3.82***
5. Used audiovisuals which are visible and/or audible	5.33	(0.62)	4.57	(0.78)	2.99**
6. Varied types of audiovisuals used.	3.07	(2.06)	2.10	(1.34)	1.52
7. When showing slides, provided sufficient light for note taking	5.45	(0.16)	4.93	(0.27)	5.50****
STUDENT INVOLVEMENT					
1. Greeted students with a bit of small talk	3.90	(1.63)	3.47	(1.81)	0.69
2. Asked questions for student consideration	3.37	(2.14)	2.80	(1.87)	0.77
3. Allowed enough time for students to think and respond	4.94	(1.18)	4.14	(1.25)	1.31
4. Used rhetorical questions to re-engage student attention	4.13	(1.63)	3.07	(1.74)	1.33
5. Encouraged students to answer difficult questions by providing cues or rephrasing	4.57	(1.13)	3.70	(1.40)	1.19
6. Asked probing questions if a student's answer was incomplete or superficial	3.33	(2.52)	4.17	(0.29)	-0.57

	High		Low		
	Mean	(SD)	Mean	(SD)	t
STUDENT INVOLVEMENT(cont.)					
7. Repeated answers when necessary so the entire class could hear	5.00	(0.71)	3.06	(1.61)	3.14*
8. Received student questions politely and when possible enthusiastically	5.58	(0.45)	4.92	(0.42)	3.80***
9. Answered students questions satisfactorily	5.54	(0.38)	4.77	(0.41)	4.75****
10. Noted and responded to signs of puzzlement, boredom, curiosity, etc.	4.20	(1.30)	2.75	(1.47)	1.91†
11. Varied the pace of the lecture to keep students alert	4.90	(0.63)	2.80	(1.19)	6.03****

† $p<.10$
* $p>.05$
** $p>.01$
*** $p>.001$
**** $p>.0001$